We claim:

1. A kit containing:

i) a scraping instrument for collecting a biological sample, comprising:

a) a proximal handle end;

b) a distal collection end; and

c) a joining portion between the handle end and the collection end;

wherein the joining portion is generally continuous in width with the handle end and the collection end on either side of the joining portion; and the joining portion allows the handle end and the collection end to be optionally detached from each other; and

wherein the collection end further comprises a peripheral edge and a depression, wherein at least some of the peripheral edge of said collection portion is serrated to allow scraping of the biological sample, and the depression allows the scraped biological sample to be collected;

ii) a storage vessel; and

iii) a stabilizing solution.

2. The kit of claim 1, wherein said collection end is spoon shaped.

3. The kit of claim 1, wherein the instrument comprises plastic.

4. The kit of claim 1, wherein the joining portion comprises a perforation.

5. The kit of claim 1, wherein the length of the instrument from about the proximal end of the handle end to the distal end of the collection end is about 3-6 inches.

6. The kit of claim 1, wherein the length of the collection end is about 1-2 inches.

7. The kit of claim 1, wherein the length and the width of the collection end allow the collection end to fit into a storage vessel.

8. The kit of claim 1, wherein the sample is comprised of epithelial cells from buccal mucosa of a subject.

9. The kit of claim 1, wherein the biological sample contains a nucleic acid.

10. The kit of claim 1, wherein the nucleic acid is selected from the group consisting of RNA and DNA.

11. The kit of claim 1, wherein the storage vessel contains a lid.

12. The kit of claim 11, wherein the lid is attached to the storage vessel.

13. An RNA collection system, comprising:

(a) a scraping instrument having a proximal handle end, a distal collection end comprising a serrated peripheral edge, and a joining portion between the handle end and the collection end, the joining portion allows the handle end and the collection end to be optionally detached from each other; and

(b) a storage vessel comprising an RNA stabilization solution.

14. The kit of claim 13, wherein the storage vessel contains a lid.

15. The kit of claim 14, wherein the lid is attached to the storage vessel.

16. A kit for collecting epithelial cells from buccal mucosa, comprising:

(a) a scraping instrument having a proximal handle end, a distal collection end comprising a serrated peripheral edge, and a joining portion between the handle end and the collection end, the joining portion allows the handle end and the collection end to be optionally detached from each other; and

(b) a storage vessel comprising an RNA stabilization solution.

17. A non-invasive method for obtaining isolated nucleic acid from mouth epithelial cells, comprising:

(a) transferring non-invasively isolated cells from a subject's mouth to a nucleic acid stabilization solution that inactivates nucleases, and

(b) extracting the nucleic acid of interest from the isolated cells, to obtain an isolated nucleic acid sample.

18. A scraping instrument for collecting a nucleic acid sample, comprising:

a) a proximal handle end;

b) a distal collection end; and

c) a joining portion between the handle end and the collection end;

wherein the joining portion is generally continuous in width with the handle end and the collection end on either side of the joining portion; and the joining portion allows the handle end and the collection end to be optionally detached from each other; and wherein the collection end further comprises a peripheral edge and a depression, wherein at least some of the peripheral edge of said collection portion is serrated to allow scraping of the nucleic acid sample, and the depression allows the scraped nucleic acid sample to be collected.

19. A method for collecting a sample, comprising the steps of:

(a) providing a scraping instrument having a proximal handle end, a distal collection end comprising a serrated peripheral edge, and a joining portion between the handle end and the collection end;

(b) providing a storage vessel comprising an RNA stabilization solution;

(c) scraping the epithelial cells from the buccal mucosa of subject's mouth with the serrated peripheral edge of the collection end;

(d) collecting the scraped epithelial cells in the collection end of the scraping instrument;

(e) transferring the scraped epithelial cells into the storage vessel; and

(f) pivoting the scraping instrument handle to cause the handle end of the instrument to detach from the collection end at the joining portion, such that the storage vessel comprises the RNA storage solution, the scraped sample, and the collection end of the scraping instrument.

20. The method of claim 17, wherein the nucleic acid is RNA.

21. The method of claim 17, wherein the cells are isolated non-invasively from the mouth by scraping with a scraping instrument.

22. The method of claim 21, wherein the scraping instrument is a plastic tool capable of collecting a large number of epithelial cells from buccal mucosa in relatively non-invasive fashion, wherein the plastic tool comprises a serrated edge to scrape off several layers of epithelial cells, and a curved surface to collect those cells.

23. The method of claim 20, wherein the sample of scraped cells in the RNA stabilization solution is stored at -15 to -25° C prior to extraction of the RNA from the sample.

24. The method of claim 23, wherein the RNA stabilization solution is RNALater RNA stabilization reagent.

25. A method for detecting the expression of a target gene(s) of interest in a sample of buccal mucosa epithelial cells, comprising:

(a) isolating a nucleic acid sample from buccal mucosa epithelial cells using the method of claim 17;

(b) contacting the isolated nucleic acid sample of step (a) with at least one nucleic acid probe which specifically hybridizes to the target gene(s) of interest; and

(c) detecting the presence of said target gene(s) of interest in the nucleic acid sample.

25. The method of claim 24, wherein the gene of interest is expressed in subjects who have lung cancer and not expressed in subjects who do not have lung cancer.

26. The method of claim 25, wherein said target gene(s) of interest is attached to a solid phase prior to performing step (b).

27. The method of claim 25, wherein the nucleic acid is RNA.

28. The method of claim 25, wherein the nucleic acid is DNA.

29. A mouth transcriptome comprising a group consisting of genes encoding ABCC1; ABHD2; AF333388.1; AGTPBP1; AIP1; AKR1B10AKR1C1; AKR1C2; AL117536.1; AL353759; ALDH3A1; ANXA3; APLP2; ARHE; ARL1; ARPC3; ASM3A; B4GALT5; BECN1; C1orf8; C20orf111; C5orf6; C6orf80; CA12; CABYR; CANX; CAP1; CCNG2; CEACAM5; CEACAM6; CED-6; CHP; CHST4; CKB; CLDN10; CNK1; COPB2; COX5A; CPNE3; CRYM; CSTA; CTGF; CYP1B1; CYP2A6; CYP4F3; DEFB1; DIAPH2; DKFZP434J214; DKFZP564K0822; DKFZP566E144; DSCR5; DSG2; EPAS1; EPOR; FKBP1A; FLJ10134; FLJ13052; FLJ130521; FLJ20359; FMO2; FTH1; GALNT1; GALNT3; GALNT7; GCLC; GCLM; GGA1; GHITM; GMDS; GNE; GPX2; GRP58; GSN; GSTM3; GSTM5; GUK1;HIG1; HIST1H2BK; HN1; HPGD; HRIHFB2122; HSPA2; IDH1; IDS; IMPA2; ITM2A; JTB; KATNB1; KDELR3; KIAA0397; KIAA0905;KLF4; KRT14; KRT15; LAMP2;LOC51186; LOC57228; LOC92482; LOC92689; LYPLA1; MAFG; ME1; MGC4342; MGLL; MT1E; MT1F; MT1G; MT1H; MT1X; MT2A; NCOR2; NKX3-1; NQO1; NUDT4; ORL1; P4HB; PEX14; PGD; PRDX1; PRDX4; PSMB5; PSMD14; PTP4A1; PTS;RAB11A;RAB2; RAB7; RAP1GA1; RNP24;

RPN2;S100A10; S100A14; S100P; SCP2; SDR1; SHARPl; SLC17A5; SLC35A3; SORD; SPINT2; SQSTM1; SRPUL; SSR4; TACSTD2; TALDO1; TARS; TCF7L1; TIAM1; TJP2; TLE1; TM4SF1; TM4SF13; TMP21; TNFSF13; TNS; TRA1; TRIM16; TXN; TXNDC5; TXNL; TXNRD1; UBE2J1; UFD1L; UGT1A10; YF13H12; and ZNF463.

30. A mouth transcriptome comprising a group consisting of genes encoding AGTPBP1; AKR1C1; AKR1C2; ALDH3A1; ANXA3; CA12; CEACAM6; CLDN10; CYP1B1; DPYSL3; FLJ13052; FTH1; GALNT3; GALNT7; GCLC; GCLM; GMDS; GPX2; HN1; HSPA2; MAFG; ME1; MGLL; MMP10; MT1F; MT1G; MT1X; NQO1; NUDT4; PGD; PRDX1; PRDX4; RAB11A; S100A10; SDR1; SRPUL; TALDO1; TARS; TCF-3; TRA1; TRIM16; and TXN.

31. A method of determining whether an individual is at increased risk of developing a lung disease, comprising:

a) taking a biological sample from the mouth of an individual exposed to an airway pollutant or at risk of being exposed to an airway pollutant; and

b) analyzing whether there is a genetic alteration in at least one gene of the mouth transcriptome genes of claim 29, wherein the presence of a genetic alteration in one or more of the mouth transcriptome genes as compared to the same at least one gene in a group of control individuals is indicative that the individual has an increased risk of developing a lung disease.

32. The method of claim 31, wherein the genetic alteration is selected from the group consisting of deviation of a gene's DNA methylation pattern and deviation of a gene's expression pattern.

33. The method of claim 32, wherein the genetic alteration is a deviation of a gene's expression pattern.

34. The method of claim 33, wherein the air pollutant is smoke from a cigarette or a cigar and the lung disease is lung cancer.

35. The method of claim 34, wherein the lung cancer is selected from adenocarcinoma, squamous cell carcinoma, small cell carcinoma, large cell carcinoma, and benign neoplasms of the lung.

36. The method of claim 34 or 35, wherein the individual is a smoker and one looks at expression of at least one gene selected from the group consisting of mouth transcriptome genes, wherein lower expression of that at least one gene in the smoker than in a control group of corresponding smokers is indicative of an increased risk of developing lung cancer.

37. The method of claim 36, wherein lower expression of at least three genes of the mouth transcriptome is indicative of an increased risk of developing lung cancer.

38. The method of claim 34 or 35, wherein the individual is a smoker and one looks at expression of at least one gene selected from the group consisting of mouth transcriptome genes, wherein higher expression of that at least one gene in the smoker than in a control group of corresponding smokers is indicative of an increased risk of developing lung cancer.

39. The method of claim 38, wherein higher expression of at least three genes selected from the group consisting of mouth transcriptome genes is indicative of an increased risk of developing lung cancer.

40. The method of claim 34 or 35, wherein the individual is a smoker and one looks at expression of at least one gene selected from the mouth transcriptomes encoding proto-oncogenes, wherein higher expression of that at least one gene in the smoker than in a control group of corresponding smokers is indicative of an increased risk of developing lung cancer.

41. The method of claim 40, wherein higher expression of at least one gene in each of the mouth transcriptome encoding proto-oncogenes is indicative of an increased risk of developing lung cancer.

42. The method of claim 34 or 35, wherein the individual is a smoker and one looks at expression of at least one gene selected from a mouth transcriptome encoding tumor suppressor genes, wherein lower expression of that at least one gene in the smoker than in a control group of corresponding smokers is indicative of an increased risk of developing lung cancer.

43. The method of claim 42, wherein lower expression of at least one gene in each of the mouth transcriptome encoding tumor suppressor genes is indicative of an increased risk of developing lung cancer.

44. A method of diagnosing predisposition of a smoker to lung disease comprising analyzing an expression pattern of one or more genes selected from the group consisting of ABCC1; ABHD2; AF333388.1; AGTPBP1; AIP1; AKR1B10AKR1C1; AKR1C2; AL117536.1; AL353759; ALDH3A1; ANXA3; APLP2; ARHE; ARL1; ARPC3; ASM3A; B4GALT5; BECN1; C1orf8; C20orf111; C5orf6; C6orf80; CA12; CABYR; CANX; CAP1; CCNG2; CEACAM5; CEACAM6; CED-6; CHP; CHST4; CKB; CLDN10; CNK1; COPB2; COX5A; CPNE3; CRYM; CSTA; CTGF; CYP1B1; CYP2A6; CYP4F3; DEFB1; DIAPH2; DKFZP434J214; DKFZP564K0822; DKFZP566E144; DSCR5; DSG2; EPAS1; EPOR; FKBP1A; FLJ10134; FLJ13052; FLJ13052l; FLJ20359; FMO2; FTH1; GALNT1; GALNT3; GALNT7; GCLC; GCLM; GGA1; GHITM; GMDS; GNE; GPX2; GRP58; GSN; GSTM3; GSTM5; GUK1;HIG1; HIST1H2BK; HN1; HPGD; HRIHFB2122; HSPA2; IDH1; IDS; IMPA2; ITM2A; JTB; KATNB1; KDELR3; KIAA0397; KIAA0905;KLF4; KRT14; KRT15; LAMP2;LOC51186; LOC57228; LOC92482; LOC92689; LYPLA1; MAFG; ME1; MGC4342; MGLL; MT1E; MT1F; MT1G; MT1H; MT1X; MT2A; NCOR2; NKX3-1; NQO1; NUDT4; ORL1; P4HB; PEX14; PGD; PRDX1; PRDX4; PSMB5; PSMD14; PTP4A1; PTS;RAB11A;RAB2; RAB7; RAP1GA1; RNP24; RPN2;S100A10; S100A14; S100P; SCP2; SDR1; SHARPl; SLC17A5; SLC35A3; SORD; SPINT2; SQSTM1; SRPUL; SSR4; TACSTD2; TALDO1; TARS; TCF7L1; TIAM1; TJP2; TLE1; TM4SF1; TM4SF13; TMP21; TNFSF13; TNS; TRA1; TRIM16; TXN; TXNDC5; TXNL; TXNRD1; UBE2J1; UFD1L; UGT1A10; YF13H12; and ZNF463 in a biological sample taken from the mouth of the smoker, wherein a divergent expression pattern of one or more of these genes as compared to the expression pattern of these genes in group of control individuals is indicative of the predisposition of the individual to lung disease.

45. A method of diagnosing predisposition of a smoker to lung disease comprising analyzing an expression pattern of one or more genes selected from the group consisting of AGTPBP1; AKR1C1; AKR1C2; ALDH3A1; ANXA3; CA12; CEACAM6; CLDN10; CYP1B1; DPYSL3; FLJ13052; FTH1; GALNT3; GALNT7; GCLC; GCLM; GMDS; GPX2; HN1; HSPA2; MAFG; ME1; MGLL; MMP10; MT1F; MT1G; MT1X; NQO1; NUDT4; PGD; PRDX1; PRDX4; RAB11A; S100A10; SDR1; SRPUL; TALDO1; TARS; TCF-3; TRA1; TRIM16; and TXN in a biological sample taken from the mouth of the smoker, wherein a divergent expression pattern of one or more of these genes as compared to the expression pattern of these genes in group of control individuals is indicative of the predisposition of the individual to lung disease.

46. A method of diagnosing predisposition of a non-smoker to lung disease comprising analyzing an expression pattern of one or more genes selected from the group consisting of outlier genes in a biological sample taken from the mouths of the non-smoker, wherein outlier genes are defined as those genes divergently expressed in the subset of smokers who develop lung cancer as compared to those smokers who do not develop lung cancer, wherein a divergent expression pattern of one or more of these genes as compared to the expression pattern of these genes in group of control individuals is indicative of the predisposition of the individual to lung disease.

47. The method of claim 45 or 46, wherein the lung disease is lung cancer.

48. The method of claim 47, wherein the lung cancer is selected from adenocarcinoma, squamous cell carcinoma, small cell carcinoma, large cell carcinoma, and benign neoplasms of the lung.

49. The method of any of claims 31-48, wherein the biological sample is a nucleic acid sample.

50. The method of claim 49, wherein the nucleic acid is RNA or DNA..

51. The method of claims 50, wherein the analysis is performed using a nucleic acid array.

52. The method of claim 50,.wherein the analysis is performed using quantitative real time PCR or mass spectrometry.